Human-in-the-loop machine learning: a state of the art

Year: 2022 | Citations: 571 | Authors: E. Mosqueira-Rey, Elena Hernández-Pereira, David Alonso-Ríos, José Bobes-Bascarán,

Abstract

Researchers are defining new types of interactions between humans and machine learning algorithms generically called human-in-the-loop machine learning. Depending on who is in control of the learning process, we can identify: active learning, in which the system remains in control; interactive machine learning, in which there is a closer interaction between users and learning systems; and machine teaching, where human domain experts have control over the learning process. Aside from control, humans can also be involved in the learning process in other ways. In curriculum learning human domain experts try to impose some structure on the examples presented to improve the learning; in explainable AI the focus is on the ability of the model to explain to humans why a given solution was chosen. This collaboration between AI models and humans should not be limited only to the learning process; if we go further, we can see other terms that arise such as Usable and Useful AI. In this paper we review the state of the art of the techniques involved in the new forms of relationship between humans and ML algorithms. Our contribution is not merely listing the different approaches, but to provide definitions clarifying confusing, varied and sometimes contradictory terms; to elucidate and determine the boundaries between the different methods; and to correlate all the techniques searching for the connections and influences between them.